Wasserstein bounds through Stein's method with bespoke derivatives

Germain Gilles and Yvik Swan

Université libre de Bruxelles

October 17, 2023


SEA of infinity
CANTORION
SET-LAND
GÖDEL'S MTS.
AXIOM-COAST
CHANNEL OF RELATIONS
LOGICA
DESERT OF THE PARADOXES
TURING'S ISLANDS
SEA of chaos
BINARY SEA
Ocean of theorems
BIT-RIVER
ALGORITHMION
COMPUTERION
Mt. C
PROGRAMMING LANGUAGES' MTS.
CITY OF PRIME NUMBERS
ABEL'S HAVEN
GALOIS BAY
POLYNOMIATH
MATRIX-HILLS
ALGEBRION
RIVER OF CRYPTOLOGY
NUMEROTHEORETICA
FERMATOR
VECTOR-WOOD
FRACTAL MTS.
LAND of the RINGS
BAY OF PERMUTATIONS
MTS. OF GROUPS
FIELD OF REAL NUMBERS
SEA OF GRAPH THEORY
COMBINATHORIATH
CAP PASCAL
BINOMIATH
GULF of PROBABILITY
Coast of approximations
NUMERIATH
OPTIMIZATION-RIVER
ISLAND of APPLICATIONS
STATISTICOR
RANDOM HILLS
INTEGRIATH
POLYHEDRON
CARTESION
Mt. π
SINUSOIDAL MTS.
EUCLIDIATH
GEOMETOR
PYTHAGORAS TRIANGLE
LOBACHEVSKI LAKE
BOLYAI LAKE
NON-EUCLIDON
ARCHIMEDIATH
PROJECTIVE PLANE
TOPOLOGIATH
TANGENT PLANE
COMPLEXIATH
EULERION
GAUSSIOR
forest of differential geometry
ANALYTHICAS
LOGARITHMIC MARSHES
Mt. MAXIMUM
MINIMUM VALLEY
CONVERGENT RIVER
NEWTON
LEIBNIZON
DERIVABILION
forest of partial derivations
SWAMP OF LIMITS
CAUCHY'S ε-CITY
CONTINUOR
SEA of functions
The Land of Middle Math

The Wasserstein distance between the laws of two rrvs X and Z is

$$W_1(X, Z) = \sup_{h \in \mathrm{Lip}(\mathbf{1})} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]|$$

where $\mathrm{Lip}(1)$ is the set of all Lipschitz functions on $\mathbb{R}$ (a.k.a. Kantorovitch or L^1 distance).

The Wasserstein distance between the laws of two rrvs X and Z is

$$W_1(X, Z) = \sup_{h \in \mathrm{Lip}(\mathbf{1})} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]|$$

where $\mathrm{Lip}(1)$ is the set of all Lipschitz functions on $\mathbb{R}$ (a.k.a. Kantorovitch or L^1 distance).

We are interested in the case where

- X is discrete:

$$\mathbb{P}^X[A] = \sum_{x_i \in A} p_i$$

with $I = \{0, \ldots, \ell\}$ or $I = \mathbb{N}$, and $x_{i-1} < x_i$ for all $i \in I \setminus \{0\}$;

The Wasserstein distance between the laws of two rrvs X and Z is

$$W_1(X, Z) = \sup_{h \in \mathrm{Lip}(\mathbf{1})} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]|$$

where $\mathrm{Lip}(1)$ is the set of all Lipschitz functions on $\mathbb{R}$ (a.k.a. Kantorovitch or L^1 distance).

We are interested in the case where

- X is discrete:
$$\mathbb{P}^X[A] = \sum_{x_i \in A} p_i$$
with $I = \{0, \dots, \ell\}$ or $I = \mathbb{N}$, and $x_{i-1} < x_i$ for all $i \in I \setminus \{0\}$;
- Z is continuous:
$$\mathbb{P}^Z[A] = \int_A q(x)dx$$
with $q : \mathbb{R} \to \mathbb{R}^+$ continuously differentiable with support $(a, b) \subseteq \mathbb{R}$.

The Wasserstein distance between the laws of two rrvs X and Z is

$$W_1(X, Z) = \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]|$$

where $\mathrm{Lip}(1)$ is the set of all Lipschitz functions on $\mathbb{R}$ (a.k.a. Kantorovitch or L^1 distance).

We are interested in the case where

- X is discrete:

$$\mathbb{P}^X[A] = \sum_{x_i \in A} p_i$$

  with $I = \{0, \ldots, \ell\}$ or $I = \mathbb{N}$, and $x_{i-1} < x_i$ for all $i \in I \setminus \{0\}$;
- Z is continuous:

$$\mathbb{P}^Z[A] = \int_A q(x) dx$$

  with $q : \mathbb{R} \to \mathbb{R}^+$ continuously differentiable with support $(a, b) \subseteq \mathbb{R}$.

We assume (i) $\mathbb{E}|Z| = \int_a^b |x| q(x) dx < \infty$ and (ii) $x_i \in \overline{(a, b)}$ for all $i \in I$.

Example (Goldstein 2007)

Let $Y_n \sim \mathrm{Bin}(n, t)$. Then with

$$X = \frac{Y_n - nt}{\sqrt{nt(1-t)}} \text{ and } Z \sim \mathcal{N}(0,1)$$

it holds that

$$W_1(X, Z) \leq \frac{1}{\sqrt{n}} \frac{t^2 + (1-t)^2}{\sqrt{t(1-t)}}.$$

Example (Goldstein 2007)

Let $Y_n \sim \mathrm{Bin}(n, t)$. Then with

$$X = \frac{Y_n - nt}{\sqrt{nt(1-t)}} \text{ and } Z \sim \mathcal{N}(0, 1)$$

it holds that

$$W_1(X, Z) \leq \frac{1}{\sqrt{n}} \frac{t^2 + (1-t)^2}{\sqrt{t(1-t)}}.$$

Example (Goldstein and Reinert 2013)

Let Y_n be distributed according to the Pólya-Eggenberger distribution with parameters $\alpha > 0, \beta > 0$, and $m \geq 1$. Then with

$$X = \frac{Y_n}{n} \text{ and } Z \sim \mathrm{Beta}(\alpha/m, \beta/m)$$

it holds that

$$W_1(X, Z) \leq \frac{C(\alpha, \beta, m)}{n}$$

with $C(\alpha, \beta, m)$ an explicit function of the parameters.

Theorem (GS, to appear)

Let $w : \mathbb{R} \to \mathbb{R}$ *be differentiable over* (a, b)*, and define* $s(x) = (q(x)w(x))'/q(x)$ *for all* $x \in (a, b)$*.*

Theorem (GS, to appear)

Let $w : \mathbb{R} \to \mathbb{R}$ *be differentiable over* (a, b), *and define* $s(x) = (q(x)w(x))'/q(x)$ *for all* $x \in (a, b)$. *Suppose furthermore that*

$$\mathbb{E}[s(X)] = 0 \quad \textit{and} \quad \mathbb{E}\left[X s(X) + w(X)\right] = 0. \tag{1}$$

Theorem (GS, to appear)

Let $w : \mathbb{R} \to \mathbb{R}$ *be differentiable over* (a, b)*, and define* $s(x) = (q(x)w(x))'/q(x)$ *for all* $x \in (a, b)$*. Suppose furthermore that*

$$\mathbb{E}[s(X)] = 0 \quad \textit{and} \quad \mathbb{E}\left[X s(X) + w(X)\right] = 0. \tag{1}$$

Then

$$W_1(X, Z) \leq C_q(w)\, \mathbb{E}\left[|\pi(X)|\delta^+(X) + |1 - \pi(X)|\delta^-(X)\right] \tag{2}$$

where

Theorem (GS, to appear)

Let $w : \mathbb{R} \to \mathbb{R}$ *be differentiable over* (a, b), *and define* $s(x) = (q(x)w(x))'/q(x)$ *for all* $x \in (a, b)$. *Suppose furthermore that*

$$\mathbb{E}[s(X)] = 0 \quad \text{and} \quad \mathbb{E}\left[Xs(X) + w(X)\right] = 0. \tag{1}$$

Then

$$W_1(X, Z) \leq C_q(w)\, \mathbb{E}\left[|\pi(X)|\delta^+(X) + |1 - \pi(X)|\delta^-(X)\right] \tag{2}$$

where

- $\delta_i^+ = \delta^+(x_i) = x_{i+1} - x_i$ *and* $\delta_i^- = \delta^-(x_i) = x_i - x_{i-1}$ *for all* $i \in I$;

Theorem (GS, to appear)

Let $w : \mathbb{R} \to \mathbb{R}$ *be differentiable over* (a, b), *and define* $s(x) = (q(x)w(x))'/q(x)$ *for all* $x \in (a, b)$. *Suppose furthermore that*

$$\mathbb{E}[s(X)] = 0 \quad \textit{and} \quad \mathbb{E}\left[Xs(X) + w(X)\right] = 0. \tag{1}$$

Then

$$W_1(X, Z) \leq C_q(w)\, \mathbb{E}\left[|\pi(X)|\delta^+(X) + |1 - \pi(X)|\delta^-(X)\right] \tag{2}$$

where

- $\delta_i^+ = \delta^+(x_i) = x_{i+1} - x_i$ *and* $\delta_i^- = \delta^-(x_i) = x_i - x_{i-1}$ *for all* $i \in I$;
- $C_q(w)$ *is an absolute constant depending only on* q *and* w;

Theorem (GS, to appear)

Let $w : \mathbb{R} \to \mathbb{R}$ *be differentiable over* (a, b)*, and define* $s(x) = (q(x)w(x))'/q(x)$ *for all* $x \in (a, b)$*. Suppose furthermore that*

$$\mathbb{E}[s(X)] = 0 \quad \text{and} \quad \mathbb{E}\left[X s(X) + w(X)\right] = 0. \tag{1}$$

Then

$$W_1(X, Z) \leq C_q(w)\, \mathbb{E}\left[|\pi(X)|\delta^+(X) + |1 - \pi(X)|\delta^-(X)\right] \tag{2}$$

where

- $\delta_i^+ = \delta^+(x_i) = x_{i+1} - x_i$ *and* $\delta_i^- = \delta^-(x_i) = x_i - x_{i-1}$ *for all* $i \in I$ *;*
- $C_q(w)$ *is an absolute constant depending only on* q *and* w*;*
- $(\pi_i)_{i\in I} := (\pi(x_i))_{i\in I}$ *is a sequence of weights given by*

$$\pi_i = \frac{1}{p_i w(x_i)} \sum_{j=0}^{i} \left((x_j - x_i)s(x_j) + w(x_j)\right) p_j. \tag{3}$$

Theorem (GS, to appear)

Let $w : \mathbb{R} \to \mathbb{R}$ *be differentiable over* (a, b), *and define* $s(x) = (q(x)w(x))'/q(x)$ *for all* $x \in (a, b)$. *Suppose furthermore that*

$$\mathbb{E}[s(X)] = 0 \quad \text{and} \quad \mathbb{E}\left[Xs(X) + w(X)\right] = 0. \tag{1}$$

Then

$$W_1(X, Z) \leq C_q(w)\,\mathbb{E}\left[|\pi(X)|\delta^+(X) + |1 - \pi(X)|\delta^-(X)\right] \tag{2}$$

where

- $\delta_i^+ = \delta^+(x_i) = x_{i+1} - x_i$ *and* $\delta_i^- = \delta^-(x_i) = x_i - x_{i-1}$ *for all* $i \in I$;
- $C_q(w)$ *is an absolute constant depending only on* q *and* w;
- $(\pi_i)_{i \in I} := (\pi(x_i))_{i \in I}$ *is a sequence of weights given by*

$$\pi_i = \frac{1}{p_i w(x_i)} \sum_{j=0}^{i} \left((x_j - x_i)s(x_j) + w(x_j)\right) p_j. \tag{3}$$

Corollary

If $x_i - x_{i-1} = \delta > 0$ *for all* $i \geq 1 \in I$ *then*

$$W_1(X, Z) \leq C_q(w)\,\delta\,\mathbb{E}[|\pi(X)| + |1 - \pi(X)|].$$

$$\pi_i = \frac{1}{p_i w(x_i)} \sum_{j=0}^{i} \left((x_j - x_i) s(x_j) + w(x_j) \right) p_j .$$

$$\pi_i = \frac{1}{p_i w(x_i)} \sum_{j=0}^{i} \left((x_j - x_i) s(x_j) + w(x_j) \right) p_j.$$

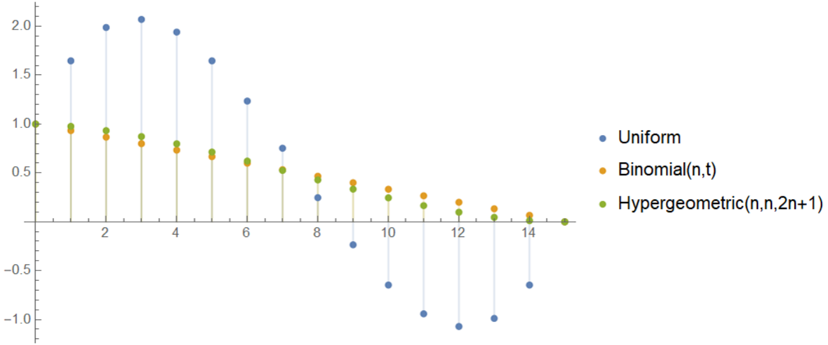

Figure: The coefficients π_i when $Z \sim N(0,1)$ ($n = 15$).

Several questions:

- Where does the bound (2) come from?
- How to choose the function w?
- What do we know about the constants $C_q(w)$?
- What does condition (1) mean?
- What can we say about the weights π from (3)?

Question 1: Where does the bound

$$W_1(X, Z) \leq C_q(w)\, \mathbb{E}\left[|\pi(X)|\delta^+(X) + |1 - \pi(X)|\delta^-(X)\right] \tag{2}$$

come from?

Question 1: Where does the bound

$$W_1(X, Z) \leq C_q(w)\,\mathbb{E}\left[|\pi(X)|\delta^+(X) + |1 - \pi(X)|\delta^-(X)\right] \tag{2}$$

come from?

Answer: A new version of Stein's method.

Definition

Stein operator for Z: Fix w some function. Define

$$\mathcal{T}_{q,w}f(x) = w(x)f'(x) + \frac{(w(x)q(x))'}{q(x)}f(x) = w(x)f'(x) + s(x)f(x).$$

Definition

Stein operator for Z: Fix w some function. Define

$$\mathcal{T}_{q,w} f(x) = w(x) f'(x) + \frac{(w(x)q(x))'}{q(x)} f(x) = w(x) f'(x) + s(x) f(x).$$

Definition

Stein operator for X: Fix $\pi = (\pi_i)_{i \in I}$ a sequence of weights. Define

$$\mathcal{T}_{p,w,\pi} f(x_i) = w(x_i) \Delta^{\pi} f(x_i) - \frac{((\Delta^{\pi})^t (wp))(x_i)}{p_i} f(x_i), \quad i \in I$$

where

$$\Delta^{\pi} f(x_i) = \pi_i \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} + (1 - \pi_i) \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}, \quad i \in I,$$

Inspired by [Yang *et al.* 2018].

Definition

Stein operator for Z: Fix w some function. Define

$$\mathcal{T}_{q,w} f(x) = w(x) f'(x) + \frac{(w(x)q(x))'}{q(x)} f(x) = w(x) f'(x) + s(x) f(x).$$

Definition

Stein operator for X: Fix $\pi = (\pi_i)_{i \in I}$ a sequence of weights. Define

$$\mathcal{T}_{p,w,\pi} f(x_i) = w(x_i) \Delta^\pi f(x_i) - \frac{((\Delta^\pi)^t (wp))(x_i)}{p_i} f(x_i), \quad i \in I$$

where

$$\Delta^\pi f(x_i) = \pi_i \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} + (1 - \pi_i) \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}, \quad i \in I,$$

Inspired by [Yang *et al.* 2018].

We suggest to use Stein's method to evaluate $W_1(X, Z)$ through comparison of $\mathcal{T}_{q,w}$ with $\mathcal{T}_{p,w,\pi}$.

Stein's method of comparison of operators in a nutshell

Let $h \in \mathrm{Lip}(1)$ and f_h be solution to

$$h(x) - \mathbb{E}[h(Z)] = \mathcal{T}_{q,w} f_h(x).$$

Stein's method of comparison of operators in a nutshell

Let $h \in \mathrm{Lip}(1)$ and f_h be solution to

$$h(x) - \mathbb{E}[h(Z)] = \mathcal{T}_{q,w} f_h(x).$$

Then $\mathbb{E}[\mathcal{T}_{p,w,\pi} f_h(X)] = 0$ for all h and

Stein's method of comparison of operators in a nutshell

Let $h \in \mathrm{Lip}(1)$ and f_h be solution to

$$h(x) - \mathbb{E}[h(Z)] = \mathcal{T}_{q,w} f_h(x).$$

Then $\mathbb{E}[\mathcal{T}_{p,w,\pi} f_h(X)] = 0$ for all h and

$$W_1(X, Z) = \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]|$$

Stein's method of comparison of operators in a nutshell

Let $h \in \mathrm{Lip}(1)$ and f_h be solution to

$$h(x) - \mathbb{E}[h(Z)] = \mathcal{T}_{q,w} f_h(x).$$

Then $\mathbb{E}[\mathcal{T}_{p,w,\pi} f_h(X)] = 0$ for all h and

$$\begin{aligned} W_1(X, Z) &= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]| \\ &= \sup_{h \in \mathrm{Lip}(1)} \left|\mathbb{E}\left[h(X) - \mathbb{E}h(Z)\right]\right| \end{aligned}$$

Stein's method of comparison of operators in a nutshell

Let $h \in \mathrm{Lip}(1)$ and f_h be solution to

$$h(x) - \mathbb{E}[h(Z)] = \mathcal{T}_{q,w} f_h(x).$$

Then $\mathbb{E}[\mathcal{T}_{p,w,\pi} f_h(X)] = 0$ for all h and

$$\begin{aligned} W_1(X, Z) &= \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}[h(X)] - \mathbb{E}[h(Z)] \right| \\ &= \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}\left[h(X) - \mathbb{E}h(Z) \right] \right| \\ &= \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}[\mathcal{T}_{q,w} f_h(X)] \right| \end{aligned}$$

Stein's method of comparison of operators in a nutshell

Let $h \in \mathrm{Lip}(1)$ and f_h be solution to

$$h(x) - \mathbb{E}[h(Z)] = \mathcal{T}_{q,w} f_h(x).$$

Then $\mathbb{E}[\mathcal{T}_{p,w,\pi} f_h(X)] = 0$ for all h and

$$\begin{aligned} W_1(X, Z) &= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]| \\ &= \sup_{h \in \mathrm{Lip}(1)} \left|\mathbb{E}\left[h(X) - \mathbb{E}h(Z)\right]\right| \\ &= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[\mathcal{T}_{q,w} f_h(X)]| \\ &= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[\mathcal{T}_{q,w} f_h(X)] - \mathbb{E}\left[\mathcal{T}_{p,w,\pi} f_h(X)\right]| \end{aligned}$$

Stein's method of comparison of operators in a nutshell

Let $h \in \mathrm{Lip}(1)$ and f_h be solution to

$$h(x) - \mathbb{E}[h(Z)] = \mathcal{T}_{q,w} f_h(x).$$

Then $\mathbb{E}[\mathcal{T}_{p,w,\pi} f_h(X)] = 0$ for all h and

$$\begin{aligned}
W_1(X, Z) &= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]| \\
&= \sup_{h \in \mathrm{Lip}(1)} \left|\mathbb{E}\left[h(X) - \mathbb{E}h(Z)\right]\right| \\
&= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[\mathcal{T}_{q,w} f_h(X)]| \\
&= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[\mathcal{T}_{q,w} f_h(X)] - \mathbb{E}\left[\mathcal{T}_{p,w,\pi} f_h(X)\right]| \\
&= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[(\mathcal{T}_{q,w} - \mathcal{T}_{p,w,\pi}) f_h(X)]|
\end{aligned}$$

Stein's method of comparison of operators in a nutshell

Let $h \in \mathrm{Lip}(1)$ and f_h be solution to

$$h(x) - \mathbb{E}[h(Z)] = \mathcal{T}_{q,w} f_h(x).$$

Then $\mathbb{E}[\mathcal{T}_{p,w,\pi} f_h(X)] = 0$ for all h and

$$\begin{aligned} W_1(X, Z) &= \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}[h(X)] - \mathbb{E}[h(Z)] \right| \\ &= \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}\left[h(X) - \mathbb{E}h(Z) \right] \right| \\ &= \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}[\mathcal{T}_{q,w} f_h(X)] \right| \\ &= \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}[\mathcal{T}_{q,w} f_h(X)] - \mathbb{E}\left[\mathcal{T}_{p,w,\pi} f_h(X) \right] \right| \\ &= \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}[(\mathcal{T}_{q,w} - \mathcal{T}_{p,w,\pi}) f_h(X)] \right| \end{aligned}$$

Hence comparing $\mathcal{T}_{q,w}$ with $\mathcal{T}_{p,w,\pi}$ directly translates into Wasserstein bounds.

This brings

$$W_1(X,Z) = \sup_{h\in\mathrm{Lip}(1)} \left| \mathbb{E}\left[w(X)\left(f_h'(X) - \Delta^{\pi} f_h(X) \right) + \left(s(X) + \frac{(\Delta^{\pi})^t(wp)}{p}(X) \right) f_h(X) \right] \right|.$$

This brings

$$W_1(X,Z) = \sup_{h\in \mathrm{Lip}(1)} \left| \mathbb{E}\left[w(X)\left(f_h'(X) - \Delta^{\pi} f_h(X) \right) + \left(s(X) + \frac{(\Delta^{\pi})^t (wp)}{p}(X) \right) f_h(X) \right] \right|.$$

From here:

- choose $\pi = \pi(p, q, w)$ to cancel out the second summand so that

$$W_1(X,Z) = \sup_{h\in \mathrm{Lip}(1)} \left| \mathbb{E}\left[w(X)\left(f_h'(X) - \Delta^{\pi} f_h(X) \right) \right] \right|$$

This brings

$$W_1(X,Z) = \sup_{h\in\mathrm{Lip}(1)} \left| \mathbb{E}\left[w(X)\left(f_h'(X) - \Delta^\pi f_h(X) \right) + \left(s(X) + \frac{(\Delta^\pi)^t (wp)}{p}(X) \right) f_h(X) \right] \right|.$$

From here:

- choose $\pi = \pi(p, q, w)$ to cancel out the second summand so that

$$W_1(X,Z) = \sup_{h\in\mathrm{Lip}(1)} \left| \mathbb{E}\left[w(X)\left(f_h'(X) - \Delta^\pi f_h(X) \right) \right] \right|$$

- Wiggling out the functions f_h from the expression leads to (2) with

$$C_q(w) = \frac{1}{2}\left(\sup_{h\in\mathrm{Lip}(1)} |(f_h' w)'|_\infty + \sup_{h\in\mathrm{Lip}(1)} |f_h'|_\infty |w'|_\infty \right).$$

Question 2: How to choose the function w and what do we know about the constants $C_q(w)$?

Question 2: How to choose the function w and what do we know about the constants $C_q(w)$?
Partial answer: The Stein kernel is a good choice of w for "simple" $Z \sim q$.

Definition

The Stein kernel for q is

$$\tau_q(x) = \frac{1}{q(x)} \int_x^b (u - \mathbb{E}[Z])q(u)du$$

a.k.a. the solution to

$$\frac{(\tau_q(x)q(x))'}{q(x)} = \mathbb{E}[Z] - x$$

for all $x \in (a, b)$.

Definition

The Stein kernel for q is

$$\tau_q(x) = \frac{1}{q(x)} \int_x^b (u - \mathbb{E}[Z])q(u)du$$

a.k.a. the solution to

$$\frac{(\tau_q(x)q(x))'}{q(x)} = \mathbb{E}[Z] - x$$

for all $x \in (a, b)$.

The constants $C_q(\tau_q)$ can (often) be computed explicitly.

Definition

The Stein kernel for q is

$$\tau_q(x) = \frac{1}{q(x)} \int_x^b (u - \mathbb{E}[Z])q(u)du$$

a.k.a. the solution to

$$\frac{(\tau_q(x)q(x))'}{q(x)} = \mathbb{E}[Z] - x$$

for all $x \in (a, b)$.

The constants $C_q(\tau_q)$ can (often) be computed explicitly.

Example

- If $Z \sim \mathcal{N}(0, 1)$ then $\tau_q(x) = 1$ and $C_q(\tau_q) = 1$.

Definition

The Stein kernel for q is

$$\tau_q(x) = \frac{1}{q(x)} \int_x^b (u - \mathbb{E}[Z])q(u)du$$

a.k.a. the solution to

$$\frac{(\tau_q(x)q(x))'}{q(x)} = \mathbb{E}[Z] - x$$

for all $x \in (a, b)$.

The constants $C_q(\tau_q)$ can (often) be computed explicitly.

Example

- If $Z \sim \mathcal{N}(0,1)$ then $\tau_q(x) = 1$ and $C_q(\tau_q) = 1$.
- If $Z \sim \mathrm{Exp}(\lambda)$ then $\tau_q(x) = x/\lambda$ and $C_q(\tau_q) = 3/2$.

Definition

The Stein kernel for q is

$$\tau_q(x) = \frac{1}{q(x)} \int_x^b (u - \mathbb{E}[Z]) q(u) du$$

a.k.a. the solution to

$$\frac{(\tau_q(x) q(x))'}{q(x)} = \mathbb{E}[Z] - x$$

for all $x \in (a, b)$.

The constants $C_q(\tau_q)$ can (often) be computed explicitly.

Example

- If $Z \sim \mathcal{N}(0,1)$ then $\tau_q(x) = 1$ and $C_q(\tau_q) = 1$.
- If $Z \sim \mathrm{Exp}(\lambda)$ then $\tau_q(x) = x/\lambda$ and $C_q(\tau_q) = 3/2$.
- If $Z \sim \mathrm{Beta}(\alpha, \beta)$ then $\tau_q(x) = x(1-x)/(\alpha+\beta)$ and an explicit bound on $C_q(\tau_q)$ is known.

Question 3: What can we say about the weights (3) and what does condition (1) mean?

Question 3: What can we say about the weights (3) and what does condition (1) mean?

Partial answer: Condition (1) is "just" a standardisation; the weights (3) are more mysterious.

Let w be given and $s = (wq)'/q$. We want π to satisfy

$$\frac{((\Delta^{\pi})^t(wp))(x_i)}{p_i} = -s(x_i), \text{ for all } i \in I \tag{4}$$

Let w be given and $s = (wq)'/q$. We want π to satisfy

$$\frac{((\Delta^{\pi})^t(wp))(x_i)}{p_i} = -s(x_i), \text{ for all } i \in I \tag{4}$$

Theorem (GS, to appear)

Recall

$$\pi_i = \frac{1}{p_i w(x_i)} \sum_{j=0}^{i} ((x_j - x_i)s(x_j) + w(x_j))\, p_j, \quad i \in I \tag{3}$$

and

$$\mathbb{E}[s(X)] = 0 \quad \textit{and} \quad \mathbb{E}\left[X s(X) + w(X)\right] = 0. \tag{1}$$

Let w be given and $s = (wq)'/q$. We want π to satisfy

$$\frac{((\Delta^{\pi})^t(wp))(x_i)}{p_i} = -s(x_i), \text{ for all } i \in I \tag{4}$$

Theorem (GS, to appear)

Recall

$$\pi_i = \frac{1}{p_i w(x_i)} \sum_{j=0}^{i} ((x_j - x_i)s(x_j) + w(x_j))\, p_j, \quad i \in I \tag{3}$$

and

$$\mathbb{E}[s(X)] = 0 \quad \textit{and} \quad \mathbb{E}[Xs(X) + w(X)] = 0. \tag{1}$$

Then

- *If I is finite the weights (3) satisfy (4) if and only if (1) holds.*

Let w be given and $s = (wq)'/q$. We want π to satisfy

$$\frac{((\Delta^{\pi})^t(wp))(x_i)}{p_i} = -s(x_i), \text{ for all } i \in I \tag{4}$$

Theorem (GS, to appear)

Recall

$$\pi_i = \frac{1}{p_i w(x_i)} \sum_{j=0}^{i} ((x_j - x_i)s(x_j) + w(x_j))\, p_j, \quad i \in I \tag{3}$$

and

$$\mathbb{E}[s(X)] = 0 \quad \textit{and} \quad \mathbb{E}\left[X s(X) + w(X)\right] = 0. \tag{1}$$

Then

- *If I is finite the weights (3) satisfy (4) if and only if (1) holds.*
- *If $I = \mathbb{N}$ the weights (3) satisfy (4), and $\pi w \in L^1(p)$ only if (1) holds.*

Let w be given and $s = (wq)'/q$. We want π to satisfy

$$\frac{((\Delta^{\pi})^t(wp))(x_i)}{p_i} = -s(x_i), \text{ for all } i \in I \tag{4}$$

Theorem (GS, to appear)

Recall

$$\pi_i = \frac{1}{p_i w(x_i)} \sum_{j=0}^{i} ((x_j - x_i)s(x_j) + w(x_j))\, p_j, \quad i \in I \tag{3}$$

and

$$\mathbb{E}[s(X)] = 0 \quad \text{and} \quad \mathbb{E}\left[Xs(X) + w(X)\right] = 0. \tag{1}$$

Then

- *If I is finite the weights (3) satisfy (4) if and only if (1) holds.*
- *If $I = \mathbb{N}$ the weights (3) satisfy (4), and $\pi w \in L^1(p)$ only if (1) holds.*

Condition (1) further simplifies if q is integrated pearson, i.e. if its Stein kernel is a 2d degree polynomial (normal, beta, gamma, student, ...), and if $w = \tau_q$ to read:

$$\mathbb{E}[X] = \mathbb{E}[Z] \text{ and } \mathrm{Var}[X] = \mathrm{Var}[Z].$$

Example

Let $Y \sim \mathrm{Bin}(n, t)$ and $X = (Y - nt)/\sqrt{nt(1-t)}$. Let $Z \sim \mathcal{N}(0, 1)$ and $w(x) = 1$. Then

$$\pi_i = 1 - i/n$$

for $i = 0, \ldots, n$. Note how $\pi_i \in [0, 1]$ for all $i = 0, \ldots, n$.

Example

Let $Y \sim \mathrm{Bin}(n, t)$ and $X = (Y - nt)/\sqrt{nt(1-t)}$. Let $Z \sim \mathcal{N}(0,1)$ and $w(x) = 1$. Then

$$\pi_i = 1 - i/n$$

for $i = 0, \ldots, n$. Note how $\pi_i \in [0,1]$ for all $i = 0, \ldots, n$.

Example

Let $Y \sim \mathrm{NBin(n,t)}$ and $X = (Y - n(1-t)/t)/\sqrt{n(1-t)/t^2}$. Let $Z \sim \mathcal{N}(0,1)$ and $w(x) = 1$. Then

$$\pi_i = 1 + i/n$$

for all $i \geq 0$. Note how $\pi_i \geq 1$ for all $i \geq 0$.

Example

Let $Y \sim \text{Bin}(n,t)$ and $X = (Y - nt)/\sqrt{nt(1-t)}$. Let $Z \sim \mathcal{N}(0,1)$ and $w(x) = 1$. Then

$$\pi_i = 1 - i/n$$

for $i = 0, \ldots, n$. Note how $\pi_i \in [0,1]$ for all $i = 0, \ldots, n$.

Example

Let $Y \sim \mathrm{NBin(n,t)}$ and $X = (Y - n(1-t)/t)/\sqrt{n(1-t)/t^2}$. Let $Z \sim \mathcal{N}(0,1)$ and $w(x) = 1$. Then

$$\pi_i = 1 + i/n$$

for all $i \geq 0$. Note how $\pi_i \geq 1$ for all $i \geq 0$.

Example

In the Polya vs beta case, we have

$$\pi_i = \frac{(i+A)(-i+n)\left(B(B+n-\sqrt{n(A+B+n)}) + A(B-n+\sqrt{n(A+B+n)})\right)}{\left(A(-i+n) + B(-i+\sqrt{n(A+B+n)})\right)\left(Bi + A(i-n+\sqrt{n(A+B+n)})\right)}.$$

for $i = 0, \ldots, n$, where $A = \alpha/m$ and $B = \beta/m$. We can show that $\pi_i \in [0,1]$.

Example

Consider the uniform measure on the set of the $\binom{n}{n/2}$ eigenvalues of the Bernoulli-Laplace Markov chain appearing in proportion to their multiplicities. Let Y be a random variable chosen from this distribution. Set

$$X = \frac{n}{2} Y + 1 \text{ and } Z \sim \mathrm{Exp}(1).$$

Fix $w(x) = x$. Then $C_q(w) = \frac{3}{2}$, $\pi_i = \frac{(2n-i)(i+1)}{2n(1+2i)} \in [0,1]$ so that

$$W_1(X, Z) \leq \frac{5}{\sqrt{n}}.$$

See also [Chatterjee and Shao 2011] for a constant equal to 12.

Example

Consider the uniform measure on the set of the $\binom{n}{n/2}$ eigenvalues of the Bernoulli-Laplace Markov chain appearing in proportion to their multiplicities. Let Y be a random variable chosen from this distribution. Set

$$X = \frac{n}{2}Y + 1 \text{ and } Z \sim \mathrm{Exp}(1).$$

Fix $w(x) = x$. Then $C_q(w) = \frac{3}{2}$, $\pi_i = \frac{(2n-i)(i+1)}{2n(1+2i)} \in [0,1]$ so that

$$W_1(X, Z) \leq \frac{5}{\sqrt{n}}.$$

See also [Chatterjee and Shao 2011] for a constant equal to 12.

Example

Let X have the stationary distribution of an $M/M/n$ queuing system with rates $\lambda, \mu > 0$ with $\lambda < \mu n$. Let Z be a continuous variable having the stationary distribution of a suitable piecewise Ornstein-Uhlenbeck process. Then [a slightly more complicated version of our theorem] gives

$$W_1\left(\sqrt{\frac{\mu}{\lambda}}(X - \frac{\lambda}{\mu}), Z\right) \leq 31\sqrt{\frac{\mu}{\lambda}}.$$

See also [Braverman *et al.* 2017] for constant equal to 205.

Example (Hypergeometric vs normal)

Let Y be the number of marked balls in a sample of r balls taken from an urn with N balls, n of which are marked; Y is a hypergeometric random variable with mean $\mu = rn/N$ and variance $\sigma^2 = nr(N-r)(N-n)/((N-1)N^2)$. Set

$$X = (Y-\mu)/\sigma \text{ and } Z \sim \mathcal{N}(0,1).$$

Fix $w = 1$. Then

$$\delta = 1/\sigma, \quad C_q(w) = 1$$

and

$$\pi_i = \frac{(n-i)(r-i)(Ni + (N-n)(N-r))}{nr(N-n)(N-r)} \in [0,1]$$

so that

$$W_1(X,Z) \leq \frac{1}{\sigma} = \frac{N\sqrt{N-1}}{\sqrt{nr(N-r)(N-n)}}$$

An example that doesn't work (for the moment): Normal approximation of intrinsic volumes.

Some facts:

- The weights can be negative, larger than 1, monotone or not, etc.

Some facts:

- The weights can be negative, larger than 1, monotone or not, etc.
- The weights π turn out to be explicit for many situations.

Some facts:

- The weights can be negative, larger than 1, monotone or not, etc.
- The weights π turn out to be explicit for many situations.
- We do not need the π to be explicit. We have identified conditions under which $(\pi_i)_{i\in I} \in [0,1]$ for all i; these conditions are not transparent and (probably) not optimal.

Some facts:

- The weights can be negative, larger than 1, monotone or not, etc.
- The weights π turn out to be explicit for many situations.
- We do not need the π to be explicit. We have identified conditions under which $(\pi_i)_{i \in I} \in [0, 1]$ for all i; these conditions are not transparent and (probably) not optimal.
- Whenever our method works, our constants are very good.

Some facts:

- The weights can be negative, larger than 1, monotone or not, etc.
- The weights π turn out to be explicit for many situations.
- We do not need the π to be explicit. We have identified conditions under which $(\pi_i)_{i \in I} \in [0, 1]$ for all i; these conditions are not transparent and (probably) not optimal.
- Whenever our method works, our constants are very good.

To do:

- More examples.

Some facts:

- The weights can be negative, larger than 1, monotone or not, etc.
- The weights π turn out to be explicit for many situations.
- We do not need the π to be explicit. We have identified conditions under which $(\pi_i)_{i \in I} \in [0, 1]$ for all i; these conditions are not transparent and (probably) not optimal.
- Whenever our method works, our constants are very good.

To do:

- More examples.
- What happens in other distances (e.g. Kolmogorov)?

Some facts:

- The weights can be negative, larger than 1, monotone or not, etc.
- The weights π turn out to be explicit for many situations.
- We do not need the π to be explicit. We have identified conditions under which $(\pi_i)_{i \in I} \in [0, 1]$ for all i; these conditions are not transparent and (probably) not optimal.
- Whenever our method works, our constants are very good.

To do:

- More examples.
- What happens in other distances (e.g. Kolmogorov)?
- When $Z \sim \mathcal{N}(0, 1)$ there seems to be a relation with ultra-log concavity of p and the behavior of π.

Some references:

- Braverman, Dai, and Feng. Stein's method for steady-state diffusion approximations: an introduction through the erlang-a and erlang-c models. Stochastic Systems, 6, 2017.
- Chatterjee, Fulman, and Röllin. Exponential approximation by Stein's method and spectral graph theory. ALEA 8, 2011.
- Döbler. Stein's method of exchangeable pairs for the beta distribution and generalizations. Electronic Journal of Probability 20, 2015.
- Goldstein. L^1 bounds in normal approximation. Annals of Probability 35, 2007.
- Goldstein and Reinert. Stein's method for the beta distribution and the Polya-Eggenberger urn. Journal of Applied Probability, 50, 2013.